Restriction theory of the Selberg sieve, with applications

par Ben GREEN et Terence TAO

RÉSUMÉ. Le crible de Selberg fournit des majorants pour certaines suites arithmétiques, comme les nombres premiers et les nombres premiers jumeaux. Nous demontrons un théorème de restriction L^2-L^p pour les majorants de ce type. Comme application immédiate, nous considérons l'estimation des sommes exponentielles sur les k-uplets premiers. Soient les entiers positifs $a_1, \ldots, a_k$ et $b_1, \ldots, b_k$. Ecrit $h(\theta) := \sum_{n \in X} e(n\theta)$, ou X est l'ensemble de tous $n \leqslant N$ tel que tous les nombres $a_1 n + b_1, \ldots, a_k n + b_k$ sont premiers. Nous obtenons les bornes supérieures pour $\|h\|_{L^p(\mathbb{T})}$, $p > 2$, qui sont (en supposant la vérité de la conjecture de Hardy et Littlewood sur les k-uplets premiers) d'ordre de magnitude correct. Une autre application est la suivante. En utilisant les théorèmes de Chen et de Roth et un «principe de transférence », nous demontrons qu'il existe une infinité de suites arithmétiques $p_1 < p_2 < p_3$ de nombres premiers, telles que chacun $p_i + 2$ est premier ou un produit de deux nombres premier.

ABSTRACT. The Selberg sieve provides majorants for certain arithmetic sequences, such as the primes and the twin primes. We prove an L^2–L^p restriction theorem for majorants of this type. An immediate application is to the estimation of exponential sums over prime k-tuples. Let $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ be positive integers. Write $h(\theta) := \sum_{n \in X} e(n\theta)$, where X is the set of all $n \leqslant N$ such that the numbers $a_1 n + b_1, \ldots, a_k n + b_k$ are all prime. We obtain upper bounds for $\|h\|_{L^p(\mathbb{T})}$, $p > 2$, which are (conditionally on the Hardy-Littlewood prime tuple conjecture) of the correct order of magnitude. As a second application we deduce from Chen's theorem, Roth's theorem, and a transference principle that there are infinitely many arithmetic progressions $p_1 < p_2 < p_3$ of primes, such that $p_i + 2$ is either a prime or a product of two primes for each $i = 1, 2, 3$.

1. Introduction and statement of results

The exponential sum over the primes[1],

$$h_N(\theta) := \sum_{p \leqslant N} e(p\theta)$$

has been very extensively studied. Here of course $e(\theta) := e^{2\pi i\theta}$. Since the time of Hardy and Littlewood such sums have been treated by dividing the circle $\mathbb{R}/\mathbb{Z}$ into the *major arcs*, which is to say points near a rational with small denominator, and the complement of this set, the *minor arcs*.

The sum $h_N(\theta)$ (or at least the very closely-related exponential sum over the von Mangoldt function) is discussed in [36, Ch. 3], for example. There it is shown how to use information concerning the distribution of primes in arithmetic progressions in order to get an asymptotic for $h_N(\theta)$ on the major arcs. On the minor arcs, $h_N(\theta)$ is shown to be rather small by an entirely different method, also described in [36, Ch. 3]. One consequence of these results is that for any $p > 2$ one has

$$(1.1) \qquad \|h_N\|_{L^p(\mathbb{T})} := \left(\int_0^1 |h_N(\theta)|^p \, d\theta\right)^{1/p} \ll_p N^{1-1/p} \log^{-1} N.$$

Disregarding the multiplicative constant, (1.1) is easily seen to be tight, this being a reflection of the prime number theorem, which in particular gives $|h_N(\theta)| \sim N/\log N$ when $|\theta| \ll 1/N$. In fact, one could estimate $\|h_N\|_p$ asymptotically.

In this paper we consider generalizations of this exponential sum in which the primes are replaced by such sets as the twin primes. More precisely, let $F = \prod_{j=1}^k (a_j n + b_j)$ be a polynomial which is product of $k \geqslant 1$ linear factors with integer coefficients. We assume that $a_j \neq 0$ and that the discriminant

$$\Delta := \prod_{1 \leqslant i < j \leqslant k} (a_i b_j - a_j b_i)$$

is non-zero, or in other words that no two linear factors are rational multiples of each other. For any such F, we consider the set

$$X = X(F) := \{n \in \mathbb{Z}^+ : F(n) \text{ is the product of } k \text{ primes }\}.$$

---

[1]The letter p will be used both for primes and for exponents in Hölder's inequality; there should be no danger of confusion. All summations over p are over the primes, and summations over other variables are over the positive integers unless otherwise indicated. Most of the notation in the introduction is very standard – in the event of further clarification being required, consultation of §2 is recommended.

For example, if $F(n) := n(n+2)$, then X is the collection of twin primes. For any $q \geqslant 1$, we introduce the quantity $\gamma(q) = \gamma(q, F)$ defined by

$$(1.2) \qquad \gamma(q) := q^{-1} |\{ n \in \mathbb{Z}/q\mathbb{Z} : (q, F(n)) = 1 \}|.$$

Observe that $0 \leqslant \gamma(q) \leqslant 1$. This number represents the proportion of residue classes modulo q which could hope to contain an infinite number of elements of X. We will impose the natural *non-degeneracy condition*

$$(1.3) \qquad \gamma(q) > 0 \text{ for all } q \geqslant 1,$$

since $X(F)$ is clearly finite if this condition fails.

Remark. Note that to verify (1.3) it suffices by the Chinese remainder theorem to verify it for prime q, and in fact one only needs to verify it for finitely many prime q, as the claim is trivial if q is larger than both k and the discriminant Δ of F. Indeed for primes p, we have

$$(1.4) \qquad \gamma(p) = 1 - k/p \text{ whenever } p > k \text{ and } p \nmid \Delta.$$

The exponential sum over the twin primes, or more generally exponential sums of the form

$$(1.5) \qquad h(\theta) = h_{N;F}(\theta) := \sum_{\substack{n \leqslant N \\ n \in X(F)}} e(n\theta),$$

have not received much attention. One very good explanation for this is that, for a fixed choice of F, the set $X(F)$ is not even known to be non-empty. Thus there is certainly no hope of obtaining an asymptotic formula for h on the major arcs, unless one is willing to obtain results on average over choices of F, as was done in [2]. On the other hand, the *Hardy-Littlewood k-tuple conjecture* asserts that for a fixed F

$$(1.6) \qquad h_{N;F}(0) = |X(F) \cap [1, N]| = (1 + o_F(1)) \mathfrak{S}_F \frac{N}{\log^k N}$$

as $N \to \infty$, where $\mathfrak{S}_F$ is the *singular series*

$$(1.7) \qquad \mathfrak{S}_F := \prod_p \frac{\gamma(p)}{(1 - \frac{1}{p})^k} \, .$$

From (1.4) we see that the infinite product in (1.7) is absolutely convergent. We will prove certain facts about $\mathfrak{S}_F$ later on, but to give the reader a feel for it we remark that if $|a_i|, |b_i| \leqslant N$ then $\mathfrak{S}_F \ll_\epsilon N^\epsilon$. Also, the L^p average of $\mathfrak{S}_F$ over any sensible family of F (for example $F(n) = n(n+h)$, $h = 1, \ldots, N$) will be absolutely bounded.

The conjecture (1.6) is still very far from being settled; however it is well known that sieve methods (either the Selberg sieve or the combinatorial

sieve) give upper bounds of the form

$$(1.8) \qquad |X(F) \cap [1, N]| \ll_k \mathfrak{S}_F (1 + o_F(1)) \frac{N}{\log^k N}$$

for fixed F (see e.g. [19, Theorem 5.3]). This upper bound only differs from (1.6) by a multiplicative constant depending on k. As an immediate consequence of this bound and Parseval's identity one may place an upper bound on $\|h_{N,F}\|_2$ which (conditionally on the k-tuple conjecture) is of the correct order of magnitude. Our first result (proven at the end of §4) asserts that we in fact can obtain good estimates for all L^p norms, $p > 2$, of these exponential sums:

Theorem 1.1. *Let F be a product of k integer linear forms having non-zero discriminant, obeying the non-degeneracy condition* (1.3) *and with coefficients a_i, b_i satisfying $|a_i|, |b_i| \leqslant N$. Let $h_{N,F}$ be the exponential sum over $X(F) \cap [1, N]$ as defined in* (1.5) *above. Then for every $2 < p < \infty$ we have the estimate*

$$\|h\|_p \ll_{p,k} \mathfrak{S}_F N^{1-1/p} (\log N)^{-k}.$$

The implied constant depends on k but is otherwise independent of F; it is explicitly computable.

Remark. In the converse direction, if the Hardy-Littlewood prime k-tuple conjecture (1.6) is true, then one easily obtains the lower bound

$$\|h\|_p \gg \mathfrak{S}_F N^{1-1/p} (\log N)^{-k},$$

at least for fixed F and N sufficiently large. This is a simple consequence of (1.6) and the observation that $h_{N;F}(\theta) \geqslant \frac{1}{2}|X(F) \cap [1, N]|$ if $|\theta| \leqslant c_k/N$ and $c_k > 0$ is sufficiently small. However, when $p = 2$, we see from Parseval's identity and the prime k-tuple conjecture (1.6) that

$$\|h\|_2 = |X(F) \cap [1, N]|^{1/2} \gg_k \mathfrak{S}_F^{1/2} N^{1/2} (\log N)^{-k/2}$$

and so Theorem 1.1 is expected to fail logarithmically at the endpoint $p = 2$.

The proof of Theorem 1.1 proceeds not by studying the set $X(F) \cap [1, N]$ directly, but rather by constructing a slightly larger object β_R, which we call an *enveloping sieve* for $X(F) \cap [1, N]$. This sieve function is considerably easier to analyze than $X(F) \cap [1, N]$ itself, and in particular enjoys very good Fourier properties. It then turns out that the only property we need on $X(F) \cap [1, N]$ itself to deduce Theorem 1.1 is that it is *majorized* by this sieve. In particular, this theorem will also hold if the set $X(F)$ is replaced by other similar sets, for instance if primes are replaced by almost primes, or with any subset of the almost primes. Combining this observation with an argument of the first author [16], we prove the following result.

Theorem 1.2. *Define a Chen prime to be a prime p such that $p + 2$ is either prime or a product of two primes. Then there are infinitely many 3-term arithmetic progressions of Chen primes.*

Remark. Such a theorem, with 3 replaced by a more general integer k, can probably be obtained by adapting the methods of [18], in which the primes were established to contain arithmetic progressions of any length k, though the explicit constants obtained by those methods are likely to be substantially worse. Our techniques here are somewhat similar in spirit to those in [18] in that they are based on establishing a *transference principle*, in this case from Roth's theorem [29] concerning progressions of length three in a dense subset of the integers, to the corresponding result concerning dense subsets of a sufficiently pseudorandom enveloping set. In [18] this transference principle was established (under quite strong pseudorandomness assumptions) by the finitary analogue of ergodic theory methods. Here, we use a much simpler harmonic analysis argument to prove the transference principle required for Theorem 1.2, and as such we require a much less stringent pseudorandomness condition (basically, we require the Fourier coefficients of the enveloping sieve to be under control).

Remark. We note that results along the lines of Theorem 1.2 can be approached using sieve methods and the Hardy-Littlewood circle method in a more classical guise. For example, Tolev [33] showed that there are infinitely many 3-term progressions $p_1 < p_2 < p_3$ of primes such that $p_i + 2$ is a product of at most r_i primes, where (r_1, r_2, r_3) can be taken to be $(5, 5, 8)$ or $(4, 5, 11)$.

The authors would like to thank the anonymous referee for a particularly attentive reading of the paper, and John Friedlander for a number of helpful remarks concerning the material in the appendix.

2. Notation

We take the opportunity to introduce here some basic notation common to the whole paper, most of which is standard in analytic number theory. Some of our notation, such as the use of the expectation symbol $\mathbb{E}$, is less standard in this field but is very convenient for certain additive problems of the type considered here. This is because there are ideas from ergodic theory bubbling just beneath the surface.

Let k be a parameter. We write $A \ll_k B$ or $A = O_k(B)$ to denote the estimate $A \leqslant C_k B$ where $C_k > 0$ is a constant depending only on k. The notation $o_k(1)$ will be reserved for a quantity which tends to zero as the variable N tends to ∞, at a rate which may depend on k (and possibly other subscripted parameters or objects).

For a statement P, we will occasionally write $\mathbf{1}_P$ to denote the indicator of P, thus $\mathbf{1}_P = 1$ if P is true and $\mathbf{1}_P = 0$ if P is false. If A is a set, we use $\mathbf{1}_A$ to denote the function $\mathbf{1}_A(x) := \mathbf{1}_{x \in A}$. We will always write $|A| = \sum_x \mathbf{1}_A(x)$ for the cardinality of A.

We write $\mathbb{Z}_q$ for the cyclic group $\mathbb{Z}/q\mathbb{Z}$, and $\mathbb{Z}_q^* := \{a \in \mathbb{Z}_q : (a, q) = 1\}$ for the multiplicatively invertible elements of $\mathbb{Z}_q$. We will often identify $\mathbb{Z}_q$ with $\{0, \ldots, q-1\}$ or with $\{1, \ldots, q\}$ in the obvious manner.

As we remarked, it is convenient to work with the language of measures and conditional expectation in much the same way as the authors did in [18]. If $f : A \to \mathbb{C}$ is a function and B is a non-empty finite subset of A, we write $\mathbb{E}_{x \in B} f(x)$ for the average value of f on B, that is to say

$$\mathbb{E}_{x \in B} f(x) := \frac{1}{|B|} \sum_{x \in B} f(x).$$

For the purposes of this paper this subscript notation is more convenient than the more traditional notation $\mathbb{E}(f(x)|x \in B)$ of probability theory. In many cases we shall just have $B = A$, in which case we abbreviate $\mathbb{E}_{x \in A} f(x)$ as $\mathbb{E}(f)$.

As remarked earlier, we write $e(\theta) = e^{2\pi i\theta}$. If m is a positive integer then we introduce the related notation e_m for the additive character on $\mathbb{Z}_m$ defined by $e_m(x) = e(x/m)$.

3. The enveloping sieve

We now begin the proof of Theorem 1.1. The strategy of the sieve method is not to study the set $X \cap [1, N]$ directly, but rather to construct an *enveloping sieve* $\beta_R : \mathbb{Z}^+ \to \mathbb{R}^+$ which is concentrated on $X \cap [1, N]$ and for which one has satisfactory control of the Fourier coefficients of β_R. We will normalize β to have average value comparable to 1; from (1.6), we would thus expect β_R to roughly be of size $\log^k N/\mathfrak{S}_F$ on $X \cap [1, N]$, and this indeed will be the case; indeed, β_R will be a *pointwise majorant* of (most of) the normalized indicator function $c_k \mathfrak{S}_F^{-1} \log^k N \mathbf{1}_{X \cap [1, N]}$.

Henceforth the polynomial F will be fixed. We shall replace the set $X = X(F)$ by "localized" versions X_q for various moduli q, defined by

$$X_q := \{n \in \mathbb{Z}_q : (q, F(n)) = 1\}.$$

This set has already appeared in (1.2). Observe that when reduced $(\operatorname{mod} q)$ all the elements in X, with at most kq exceptions, lie in X_q, the exceptions arising from the cases when $a_j n + b_j \in \{1, \ldots, q\}$ for some $1 \leqslant j \leqslant k$.

We can now describe the important properties of enveloping sieve, which is intended to majorize the set

$$X_{R!} = \{n \in \mathbb{Z} : (d, F(n)) = 1 \text{ for all } 1 \leqslant d \leqslant R\}$$

where R is a large integer (the sieve parameter) to be chosen later. Eventually we will set $R := \lfloor N^{1/20} \rfloor$.

Proposition 3.1 (The enveloping sieve)*. Let F be the product of k integer linear forms with non-zero discriminant, obeying the non-degeneracy condition (1.3) and with coefficients a_i, b_i satisfying $|a_i|, |b_i| \leqslant N$. Let $R \leqslant N$ be a large integer. Then there is a non-negative function $\beta := \beta_R : \mathbb{Z} \to \mathbb{R}^+$ with the following properties:*

(i) *(Majorant property) We have*

$$(3.1) \qquad \beta(n) \gg_k \mathfrak{S}_F^{-1} \log^k R \mathbf{1}_{X_{R!}}(n)$$

for all integers n. In particular, $\beta(n)$ is non-negative.

(ii) *(Crude upper bound) We have*

$$(3.2) \qquad \beta(n) \ll_{k,\epsilon} N^\epsilon$$

for all $0 < n \leqslant N$ and $\epsilon > 0$.

(iii) *(Fourier expansion) We have*

$$(3.3) \qquad \beta(n) = \sum_{q \leqslant R^2} \sum_{a \in \mathbb{Z}_q^*} w(a/q) e_q(-an),$$

where $w(a/q) = w_R(a/q)$ obeys the bound

$$(3.4) \qquad |w(a/q)| \ll_{k,\epsilon} q^{\epsilon - 1}$$

for all $q \leqslant R^2$ and $a \in \mathbb{Z}_q^*$. Also we have $w(0) = w(1) = 1$.

(iv) *(Fourier vanishing properties) Let $q \leqslant R^2$ and $a \in \mathbb{Z}_q^*$. If q is not square-free, then $w(a/q) = 0$. Similarly, if $\gamma(q) = 1$ and $q > 1$, then $w(a/q) = 0$.*

Remark. The enveloping sieve β_R is essentially a normalised version of the one used by Selberg [31] in obtaining the upper bound (1.8). It will be constructed quite explicitly in §7, where a full proof of Proposition 3.1 will also be supplied. One does not need to know the exact construction in order to use Proposition 3.1 in applications. In §8 we offer some remarks which explain the phrase "enveloping sieve", as well as a comparison of β_R (in the case $F(n) = n$) and a somewhat different majorant for the primes which we used in [18].

4. The Hardy-Littlewood majorant property for the enveloping sieve

In Proposition 3.1 we constructed a majorant β_R for the set $X_{R!}$ which enjoyed good control on the Fourier coefficients; indeed, from (3.4) it is easy to obtain good L^p control on the Fourier coefficients of β_R for $p > 2$, although the bound in (3.4) causes unavoidable logarithmic losses at the endpoint $p = 2$. When $p = 4, 6, 8, \ldots$ one can pass from this to good control on h by Parseval's formula, and hence by interpolation one can obtain Theorem 1.1 in the case $p \geqslant 4$. However this simple argument does not seem to suffice in the more interesting region[2] $2 < p < 4$ because of the logarithmic failure of the estimate at $p = 2$ remarked on in the introduction. In this region $2 < p < 4$ (or more generally for $p \notin 2\mathbb{N}$) there are no good monotonicity properties in L^p of the Fourier transform to exploit; see [24, p144] for a simple example where monotonicity in this sense breaks down, and [1, 17, 23] for further discussion.

On the other hand, all known examples where monotonicity breaks down are rather pathological. When the majorant β_R enjoys additional Fourier or geometric properties it is often possible to recover estimates of Hardy-Littlewood majorant type. This is known as the *restriction phenomenon* and has been intensively studied in harmonic analysis; see for instance [32] for a recent survey of this theory.

The application of ideas from restriction theory to number theory was initiated by Bourgain in the papers [3, 4] (among others). In [3], which is of the most relevance to us, it is shown how to obtain bounds for $\|\hat{f}\|_p$, where f is a function supported on the primes. Such a bound was also obtained in a paper of the first author [16], using the Brun sieve and somewhat precise versions of the prime number theorem for arithmetic progressions. Our arguments here, while certainly in a similar spirit, differ from the previous arguments in that we rely purely on the properties of the majorant given by Proposition 3.1; in particular we do not need any type of control of the Riemann ζ- function or of related objects.

Proposition 4.1 gives a restriction estimate for β in the setting of the finite abelian group $\mathbb{Z}_N$. The *dual* form of this estimate, which is Proposition 4.2, will be of more interest for applications (indeed, Theorem 1.1 will be an almost immediate consequence of it). As in many arguments in restriction theory, the argument is based in spirit on the Tomas-Stein method, first used in [34].

---

[2]For the application to finding progressions of length three, we in fact need this estimate for some $2 < p < 3$.

Proposition 4.1 (L^q Restriction estimate for β)**.** *Let R, N be large numbers such that $1 \ll R \ll N^{1/10}$ and let k, F, β_R be as in Proposition 3.1. Let $f : \mathbb{Z}_N \to \mathbb{C}$ be arbitrary. Then for every $1 < q < 2$ we have*

$$(4.1) \qquad \mathbb{E}_{1 \leqslant n \leqslant N} |\sum_{b \in \mathbb{Z}_N} f(b)e_N(bn)|^2 \beta_R(n) \ll_{q,k} \left(\sum_{b \in \mathbb{Z}_N} |f(b)|^q\right)^{2/q}.$$

Proof. We will first examine the left-hand side of (4.1) in the special case when f is supported on a set $B \subseteq \mathbb{Z}_N$ and $\|f\|_\infty \leqslant 1$. We assert that for every $\epsilon > 0$ we have

$$(4.2) \qquad \mathbb{E}_{1 \leqslant n \leqslant N} |\sum_{b \in \mathbb{Z}_N} f(b)e_N(bn)|^2 \beta_R(n) \ll_{\epsilon,k} |B|^{1+\epsilon}.$$

This statement, though at first sight weaker than (4.1), is actually equivalent to it. To deduce (4.1) from (4.2), normalise so that

$$(4.3) \qquad \left(\sum_{b \in \mathbb{Z}_N} |f(b)|^q\right)^{2/q} = 1.$$

In particular this implies that $|f(b)| \leqslant 1$ for all $b \in \mathbb{Z}_N$. If we then define the sets $B_j \subset \mathbb{Z}_N$ for all $j \geqslant 0$ as $B_j := \{b \in \mathbb{Z}_N : 2^{-j-1} < |f(b)| \leqslant 2^{-j}\}$, then we have

$$\sum_{b \in \mathbb{Z}_N} f(b)e_N(bn) = \sum_{j \geqslant 0} \sum_{b \in B_j} f(b)e_N(bn).$$

Also, from (4.3) we have $|B_j| \leqslant 2^{(j+1)q}$. Applying (4.2) to B_j (and with $f(b)$ replaced by $2^j f(b)$) we have

$$\left(\mathbb{E}_{1 \leqslant n \leqslant N} |\sum_{b \in B_j} f(b)e_N(bn)|^2 \beta_R(n)\right)^{1/2} \quad \ll_{q,\epsilon,k} \quad 2^{-j} |B_j|^{(1+\epsilon)/2}$$

$$\leqslant \quad 2^{((1+\epsilon)q/2-1)j}$$

for any $\epsilon > 0$. If we choose ϵ sufficiently small, then the exponent $(1+\epsilon)q/2 - 1$ is negative since $q < 2$. Summing this in j using the triangle inequality in $L^2_{\beta,N}$, defined as the space of complex sequences $(a_n)_{n=1}^N$ together with the inner product

$$\langle (a_n), (b_n) \rangle = \mathbb{E}_{1 \leqslant n \leqslant N} a_n \overline{b_n} \beta_R(n),$$

we see that (4.1) indeed follows (since $q < 2$).

It remains to prove (4.2) which, recall, was to be demonstrated on the assumption that $\|f\|_\infty \leqslant 1$ and that f is supported on B. When B is large, say $|B| \geqslant R^\epsilon$, we can use the crude estimate $\beta_R(n) \ll_{k,\epsilon} R^{\epsilon^2}$ (for instance) from (3.2), combined with Parseval's identity to conclude the bound

$$\mathbb{E}_{1 \leqslant n \leqslant N} |\sum_{b \in B} f(b)e_N(bn)|^2 \beta_R(n) \ll_{k,\epsilon} R^{\epsilon^2} |B| \leqslant |B|^{1+\epsilon}.$$

Thus we may take $|B| \leqslant R^\epsilon$.

We will essentially be doing Fourier analysis on $\mathbb{Z}_N$, but exponentials such as $e_q(a)$, which arise from the Fourier expansion (3.3), will also appear. To facilitate this somewhat awkward juxtaposition it will be convenient to introduce a smooth Fourier cutoff. Let $\psi : \mathbb{R} \to \mathbb{R}^+$, $\|\psi\|_\infty \leqslant 1$, be a non-negative even bump function supported on the interval $[-1/10, 1/10]$ whose Fourier transform $\widehat{\psi}(y) := \int_{\mathbb{R}} \psi(x) e(xy)\, dx$ is non-negative everywhere and bounded away from zero when $y \in [-1, 1]$. Such a function can easily be constructed, for instance by convolving a non-negative even bump function supported on $[-1/20, 1/20]$ with itself. Then we have

$$\mathbb{E}_{1 \leqslant n \leqslant N} |\sum_{b \in B} f(b) e_N(bn)|^2 \beta_R(n) \ll \frac{1}{N} \sum_{n \in \mathbb{Z}} |\sum_{b \in B} f(b) e_N(bn)|^2 \beta_R(n) \widehat{\psi}(n/N).$$

We expand the right-hand side using the Fourier expansion of β_R, (3.3), to obtain

$$\frac{1}{N} \sum_{n \in \mathbb{Z}} \sum_{q \leqslant R^2} \sum_{a \in \mathbb{Z}_q^*} \sum_{b,b' \in B} f(b)\overline{f(b')} w(a/q) e_N(bn) e_N(-b'n) e_q(-an) \widehat{\psi}(n/N).$$

Now the Poisson summation formula implies that

$$\frac{1}{N} \sum_{n \in \mathbb{Z}} e_N(bn) e_N(-b'n) e_q(-an) \widehat{\psi}(n/N) = \sum_{m \in \mathbb{Z}} \psi\left(N(m + \frac{b - b'}{N} - \frac{a}{q})\right),$$

and hence by construction of ψ

$$\left| \frac{1}{N} \sum_{n \in \mathbb{Z}} e_N(bn) e_N(-b'n) e_q(-an) \widehat{\psi}(n/N) \right| \leqslant \mathbf{1}_{\|\frac{b-b'}{N} - \frac{a}{q}\| \leqslant 1/N},$$

where $\|x\|$ denotes the distance of x to the nearest integer. Applying this and (3.4), and bounding the coefficients $f(b)$, $\overline{f(b')}$ by one, our task thus reduces to proving that

$$(4.4) \qquad \sum_{q \leqslant R^2} \sum_{a \in \mathbb{Z}_q^*} \sum_{\substack{b,b' \in B \\ \|\frac{b-b'}{N} - \frac{a}{q}\| \leqslant 1/N}} q^{\epsilon/2 - 1} \ll_\epsilon |B|^{1+\epsilon}.$$

Let us first dispose of the contribution of the large q, for which $q \geqslant |B|$, where it is possible to estimate $q^{\epsilon/2-1}$ by $|B|^{\epsilon/2-1}$. Observe that for any two distinct fractions a/q, a'/q' with $q, q' \leqslant R^2$ and $a \in \mathbb{Z}_q^*$, $a' \in \mathbb{Z}_{q'}^*$, we have

$$\left\| \frac{a}{q} - \frac{a'}{q'} \right\| \geqslant \frac{1}{qq'} \geqslant \frac{1}{R^4} \geqslant \frac{2}{N}$$

since $N \gg R^{10}$. This shows that for each fixed b, b' there is at most one fraction $\frac{a}{q}$ which contributes to the sum (4.4). Thus the contribution of the

large q is at most $|B|^2|B|^{\epsilon/2-1}$, which is acceptable. It will thus suffice to estimate the sum over small q, that is to say those q for which $q \leqslant |B|$, and to that end it is enough to prove that

$$(4.5) \qquad \sum_{q \leqslant |B|} \sum_{a \in \mathbb{Z}_q^*} \sum_{\substack{b,b' \in B \\ \|\frac{b-b'}{N} - \frac{a}{q}\| \leqslant 1/N}} q^{-1} \ll_\epsilon |B|^{1+\epsilon/2}.$$

We estimate the left-hand side by

$$c \sum_{q \leqslant |B|} \sum_{a \in \mathbb{Z}_q} \sum_{b,b' \in B} q^{-1} \sum_{m \in \mathbb{Z}} \widehat{\psi}\left(N(m - \frac{b-b'}{N} + \frac{a}{q})\right),$$

where $c = \max_{y \in [-1,1]} |\widehat{\psi}(y)|^{-1}$, which by the Poisson summation formula is

$$cN^{-1} \sum_{q \leqslant |B|} \sum_{a \in \mathbb{Z}_q} \sum_{b,b' \in B} q^{-1} \sum_{n \in \mathbb{Z}} \psi(n/N) e_N((b-b')n) e_q(-an).$$

Performing the summation over a, this becomes

$$cN^{-1} \sum_{\substack{q \leqslant |B| \\ q|n}} \sum_{b,b' \in B} \sum_{n \in \mathbb{Z}} \psi(n/N) e_N((b-b')n)$$

which we can rearrange as

$$(4.6) \qquad cN^{-1} \sum_{n \in \mathbb{Z}} \psi(n/N) \left| \sum_{b \in B} e_N(bn) \right|^2 d_{|B|}(n),$$

where

$$d_{|B|}(n) := \sum_{\substack{q \leqslant |B| \\ q|n}} 1.$$

The $n = 0$ term of this sum may be bounded, very crudely, by $O(N^{-1}|B|^3)$. Since $|B| \leqslant R^\epsilon \ll N^{1/10}$, this term is certainly at most $O(|B|)$ and hence does not make an important contribution to (4.5). Let us thus discard the $n = 0$ term, and bound the remaining terms using Hölder's inequality. Since $\psi(n/N)$ is supported in the region $|n| \leqslant N/2$, we thus obtain

$$(4.6) \ll \left(\mathbb{E}_{0<|n|\leqslant N/2} |\sum_{b \in B} e_N(bn)|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} \left(\mathbb{E}_{0<|n|\leqslant N/2} d_{|B|}(n)^{\frac{2+\epsilon}{\epsilon}}\right)^{\frac{\epsilon}{2+\epsilon}}.$$

Parseval and the crude bound $|\sum_{b \in B} e_N(bn)| \leqslant |B|$ yield

$$\left(\mathbb{E}_{0<|n|\leqslant N/2} |\sum_{b \in B} e_N(bn)|^{2+\epsilon}\right)^{\frac{2}{2+\epsilon}} \leqslant (|B|^{1+\epsilon})^{\frac{2}{2+\epsilon}} = |B|^{1+\frac{\epsilon}{2+\epsilon}}.$$

On the other hand, from standard moment estimates for the restricted divisor function $d_{|B|}$ (see for instance [3, 30]) we have

$$\mathbb{E}_{0<|n|\leqslant N/2}d_{|B|}(n)^m \ll_{\epsilon,m} |B|^\epsilon$$

for any $m \geqslant 1$ (in fact one can replace $|B|^\epsilon$ with $\log^{2^m-1}|B|$ when m is an integer - see [30]). Applying this with $m := \frac{2+\epsilon}{\epsilon}$ and working back, one gets (4.6) $\ll_\epsilon |B|^{1+\epsilon/2}$, which in turn implies (4.5) and (4.4), and hence the proposition. $\square$

As promised, we now give what is essentially a dual form of Proposition 4.1.

Proposition 4.2 (L^p extension estimate for β)**.** *Let R, N be large numbers such that $1 \ll R \ll N^{1/10}$ and let k, F, β_R be as in Proposition 3.1. Let $\{a_n\}_{n\leqslant N}$ be an arbitrary sequence of complex numbers, and let $p > 2$ be a fixed real number. Then we have*

$$(4.7) \quad \Big(\sum_{b\in\mathbb{Z}_N}|\mathbb{E}_{1\leqslant n\leqslant N}a_n\beta_R(n)e_N(-bn)|^p\Big)^{1/p} \ll_{p,k} \big(\mathbb{E}_{1\leqslant n\leqslant N}|a_n|^2\beta_R(n)\big)^{1/2}$$

and

$$(4.8) \quad \|\mathbb{E}_{1\leqslant n\leqslant N}a_n\beta_R(n)e(n\theta)\|_{L^p(\mathbb{T})} \ll_{p,k} N^{-1/p}\big(\mathbb{E}_{1\leqslant n\leqslant N}|a_n|^2\beta_R(n)\big)^{1/2}.$$

Proof. The first claim follows from Proposition 4.1 and a duality argument. We will expand on this point somewhat using the language of operators, since this will afford us an opportunity to at least partially explain the terms *restriction* and *extension*. Everything we say could also be phrased in terms of inequalities on exponential sums. Recall that $L^2_{\beta,N}$ is the space of complex sequences $(a_n)_{n=1}^N$ together with the inner product

$$\langle(a_n),(b_n)\rangle = \mathbb{E}_{1\leqslant n\leqslant N}a_n\overline{b_n}\beta_R(n).$$

Consider also the space of complex-valued functions on $\mathbb{Z}_N$ together with the l^q norms

$$\|f\|_{l^q(\mathbb{Z}_N)} := \Big(\sum_{x\in\mathbb{Z}_N}|f(x)|^q\Big)^{1/q}$$

and the inner product $\langle f,g\rangle := \sum_{x\in\mathbb{Z}_N}f(x)\overline{g(x)}$. Here, $\{1,\ldots,N\}$ may perhaps be thought of as a finite version of $\mathbb{Z}$, whilst $\mathbb{Z}_N$ plays the rôle of a discretised circle $\mathbb{T}$. Now define the *restriction* map $T : l^q(\mathbb{Z}_N) \to L^2_{\beta,N}$ by

$$f \mapsto \Big(\sum_{x\in\mathbb{Z}_N}f(x)e_N(nx)\Big)_{n=1}^N$$

(the reason for the name is that T can be thought of as the restriction of the discrete Fourier transform on $\mathbb{Z}_N$ to the weight β). One can check that

the adjoint of T is the *extension* map $T^* : L^2_{\beta,N} \to l^p(\mathbb{Z}_N)$ defined by

$$(a_n)_{n=1}^N \mapsto \mathbb{E}_{1 \leqslant n \leqslant N} a_n \beta_R(n) e_N(-nx).$$

Here, of course, p is the dual exponent to q defined by the relation $p^{-1} + q^{-1} = 1$; note, however, that all of the spaces $l^r(\mathbb{Z}_N)$ are equivalent as vector spaces.

Now one can check that (4.1), the main result of Proposition 4.1, is equivalent to

$$\|T\|_{l^q(\mathbb{Z}_N) \to L^2_{\beta,N}} \ll_{p,k} 1.$$

On the other hand (4.7) is equivalent to the statement

$$\|T^*\|_{L^2_{\beta,N} \to l^p(\mathbb{Z}_N)} \ll_{p,k} 1.$$

By the duality principle for operator norms one sees that these two statements are completely equivalent, and so (4.7) is an immediate deduction from Proposition 4.1.

To prove (4.8), observe that for any $0 \leqslant \theta < 1$ we can obtain the estimate

$$\sum_{b \in \mathbb{Z}_N} |\mathbb{E}_{1 \leqslant n \leqslant N} a_n \beta_R(n) e((b + \theta)n/N)|^p \ll_{p,k} \left(\mathbb{E}_{1 \leqslant n \leqslant N} |a_n|^2 \beta_R(n) \right)^{p/2}$$

from (4.7) simply by multiplying each coefficient a_n by $e_N(\theta n)$. Integrating this in θ from 0 to 1 gives (4.8). $\qquad \square$

Remark. A result of Marcinkiewicz [38] allows one to obtain (4.7) from (4.8); see [16] for further discussion.

The utility of Proposition 4.2 is markedly increased by the next lemma.

Lemma 4.1. *Suppose that $R \leqslant \sqrt{N}$. Then*

$$\mathbb{E}_{1 \leqslant n \leqslant N} \beta_R(n) \ll 1.$$

Proof. We make further use of the smooth cutoff function ψ. We have

$$\mathbb{E}_{1 \leqslant n \leqslant N} \beta_R(n) \ll \frac{1}{N} \sum_{n \in \mathbb{Z}} \hat{\psi}(n/N) \beta_R(n).$$

Substituting in the Fourier expansion (3.3) of β_R and using the Poisson summation formula, we have

$$N^{-1} \sum_{n \in \mathbb{Z}} \hat{\psi}(n/N) \beta_R(n) = \sum_{q \leqslant R^2} \sum_{a \in \mathbb{Z}_q^*} w(a/q) \sum_{m \in \mathbb{Z}} \hat{\psi}\left(N(m - \frac{a}{q})\right).$$

Observe that since $R^2 \leqslant N$, the only time that $\hat{\psi}(N(m - \frac{a}{q}))$ is non-zero is when $m = a = 0$ and $q = 1$. Thus, by (3.4), we obtain

$$\frac{1}{N} \sum_{n \in \mathbb{Z}} \widehat{\psi}(n/N)\beta_R(n) = w(0)\psi(0) \ll 1,$$

and the claim follows. $\qquad \square$

Observe that the last lemma, combined with Proposition 3.1(i), implies the classical sieve estimate (1.8). With these preliminaries in place we can now quickly conclude the proof of Theorem 1.1.

Proof of Theorem 1.1. Fix $R := \lfloor N^{1/20} \rfloor$ (for instance). Without loss of generality we may assume N to be sufficiently large depending on k and p. We now apply Proposition 4.2 to the sequence $a_n := \mathbf{1}_{X \cap X_{R!}}(n)/\beta(n)$. From (3.1), (4.8) and Lemma 4.1 we easily obtain

$$\Big\| \sum_{X \cap X_{R!} \cap [1,N]} e(n\theta) \Big\|_{L^p(\mathbb{T})} \ll_{p,k} \frac{\mathfrak{S}_F}{\log^k R} N^{1-1/p}.$$

Since all but at most kR elements of X lie in $X_{R!}$, the claim now follows since $R \sim N^{1/20}$ and N is assumed sufficiently large. $\qquad \square$

Remark. A straightforward modification of the above argument also gives the estimate

$$\int_0^1 | \sum_{n \leqslant N} (\prod_{j=1}^k \Lambda(a_j n + b_j))e(n\theta)|^p \, d\theta \ll_{p,k} \mathfrak{S}_F^p N^{p-1},$$

where Λ is the von Mangoldt function.

5. A transference principle for 3-term arithmetic progressions

There is a well-known theorem of Roth [29], partially answering a question of Erdős and Turán, which states that any subset of $\{1, \dots, N\}$ with size at least $cN/\log\log N$ contains an arithmetic progression of length 3. In this section we will prove a result in this spirit for sets which are majorised by a "pseudorandom measure". In §6 we will apply this result to deduce that there are infinitely many 3-term progressions of Chen primes (primes p such that $p + 2$ is prime or a product of two primes).

We refer to the result (Proposition 5.1) as a transference principle because, rather than prove it from first principles, we deduce it from Roth's theorem. For the reader familiar with [18], we remark that the transference principle we prove here is shown to hold under harmonic analysis conditions which should be regarded as much less stringent than the combinatorial conditions of that paper. However, our arguments here only apply to progressions of

length 3. Indeed, it is by now well-understood that traditional harmonic analysis arguments cannot suffice to deal with progressions of length 4 or more (see [14] for a further discussion).

We begin by recalling a formulation of Roth's theorem [29] due to Varnavides [35].

Theorem 5.1. [29, 35] *Let $0 < \delta \leqslant 1$ be a positive numbers, and let N be a large prime parameter. Suppose that $f : \mathbb{Z}_N \to \mathbb{R}^+$ is a function satisfying the uniform bounds*

$$0 \leqslant f(n) \leqslant 1 \text{ for all } n \in \mathbb{Z}_N$$

and the mean property

$$\mathbb{E}(f) \geqslant \delta.$$

Then we have

(5.1) $$\mathbb{E}_{n,d \in \mathbb{Z}_N} f(n)f(n+d)f(n+2d) \geqslant c(\delta)$$

for some constant $c(\delta) > 0$ depending only on δ.

Remark. By combining Bourgain's refinement [5] of Roth's theorem with the argument of Varnavides [35] one can obtain the more explicit value $c(\delta) = \exp(-C\delta^{-2} \log(1 + \frac{1}{\delta}))$ for $c(\delta)$; see also [16] for further discussion. However, our arguments will not be sensitive to the exact value of $c(\delta)$.

We now generalise the above proposition to the case where f is not bounded by 1, but is instead bounded by some suitably pseudorandom function.

Definition. We shall normalize the Fourier transform $\widehat{f} : \mathbb{Z}_N \to \mathbb{C}$ of a function $f : \mathbb{Z}_N \to \mathbb{C}$ according to the formula

$$\widehat{f}(a) := \mathbb{E}_{n \in \mathbb{Z}_N} f(n) e_N(an) \text{ for all } a \in \mathbb{Z}_N,$$

and recall the *Fourier inversion formula*

(5.2) $$f(n) = \sum_{a \in \mathbb{Z}_N} \widehat{f}(a) e_N(-an).$$

With this normalisation it is natural to measure the Fourier transform $\widehat{f}$ using the l^p norms

$$\|\widehat{f}\|_p^p := \sum_{a \in \mathbb{Z}_N} |\widehat{f}(a)|^p$$

for $1 < p < \infty$, with the usual convention that $\|\widehat{f}\|_\infty := \sup_{a \in \mathbb{Z}_N} |\widehat{f}(a)|$.

Proposition 5.1. *Let N be a large prime parameter. Suppose that $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ and $f : \mathbb{Z}_N \to \mathbb{R}^+$ are non-negative functions satisfying the majorization condition*

(5.3) $$0 \leqslant f(n) \leqslant \nu(n) \text{ for all } n \in \mathbb{Z}_N$$

and the mean condition

$$(5.4) \qquad \qquad \mathbb{E}_{n \in \mathbb{Z}_N} f(n) \geqslant \delta$$

for some $0 < \delta \leqslant 1$. Suppose that ν and f also obey the pseudorandomness conditions

$$(5.5) \qquad \qquad |\widehat{\nu}(a) - \delta_{a,0}| \leqslant \eta \text{ for all } a \in \mathbb{Z}_N$$

where $\delta_{a,0}$ is the Kronecker delta, and

$$(5.6) \qquad \qquad \|\widehat{f}\|_q \leqslant M$$

for some $\eta, M > 0$ and some $2 < q < 3$. Then we have

$$(5.7) \qquad \mathbb{E}_{n,d \in \mathbb{Z}_N} f(n) f(n+d) f(n+2d) \geqslant \tfrac{1}{2} c(\delta) - O_{\delta,M,q}(\eta)$$

where $c(\delta) > 0$ is the same constant that appears in Theorem 5.1.

Remark. Observe that the $\nu \equiv 1$, $\eta = 0$ case of this proposition is basically Theorem 5.1, with an (inconsequential) loss of $\frac{1}{2}$ on the right-hand side of (5.1). This theorem should be compared with [18, Theorem 3.5]. In practice, the condition (5.6) seems to be most easily verified by first establishing a Hardy-Littlewood majorant property such as Proposition 4.2 for *all* functions f majorized by ν, not just the specific function f of interest.

Proof. We repeat the arguments of [16, §6]. Let $0 < \epsilon \ll 1$ be a small parameter to be chosen later, and let

$$(5.8) \qquad \qquad \Omega := \{a \in \mathbb{Z}_N : |\widehat{f}(a)| \geqslant \epsilon\}$$

be the set of large Fourier coefficents of f. Write $B := B(\Omega, \epsilon)$ for the *Bohr set*

$$(5.9) \qquad B := B(\Omega, \epsilon) := \{m \in \mathbb{Z}_N : |1 - e_N(am)| \leqslant \epsilon \text{ for all } a \in \Omega\}.$$

We shall decompose $f = f_1 + f_2$, where f_1 is the *anti-uniform* component of f, defined by

$$(5.10) \qquad \qquad f_1(n) := \mathbb{E}_{m_1,m_2 \in B} f(n + m_1 - m_2),$$

and $f_2 := f - f_1$ will be the *uniform* component. We can thus split the left-hand side of (5.7) into eight components

$$(5.11) \qquad \qquad \mathbb{E}_{n,d \in \mathbb{Z}_N} f_i(n) f_j(n+d) f_k(n+2d)$$

where $i, j, k \in \{1, 2\}$.

Let us first consider the term (5.11) corresponding to $(i, j, k) = (1, 1, 1)$, which will be the dominant term. The pseudorandomness condition (5.5)

can be used to obtain uniform bounds on f_1. Indeed using that together with (5.2), (5.3) and Parseval's identity we obtain

$$
\begin{aligned}
|f_1(n)| &= |\mathbb{E}_{m_1,m_2\in B} f(n + m_1 - m_2)| \\
&\leqslant \mathbb{E}_{m_1,m_2\in B} \nu(n + m_1 - m_2) \\
&= |\mathbb{E}_{m_1,m_2\in B} \sum_{a\in\mathbb{Z}_N} \widehat{\nu}(a) e_N(-an - am_1 + am_2)| \\
&= |\sum_{a\in\mathbb{Z}_N} \hat{\nu}(a) e_N(-an)|\mathbb{E}_{m\in B} e_N(-am)|^2| \\
&\leqslant \sum_{a\in\mathbb{Z}_N} |\hat{\nu}(a)||\mathbb{E}_{m\in B} e_N(-am)|^2 \\
&\leqslant 1 + \eta \sum_{a\in\mathbb{Z}_N} |\mathbb{E}_{m\in B} e_N(-am)|^2 \\
&= 1 + \eta N/|B|.
\end{aligned}
$$

Note, however, that by the pigeonhole principle we have

$$
|B| \geqslant (c\epsilon)^{|\Omega|} N
$$

for some absolute constant $c > 0$, while from (5.6) and Chebyshev's inequality we have

$$
|\Omega| \leqslant (M/\epsilon)^q.
$$

Finally, f_1 is manifestly non-negative by (5.10) and (5.3). We thus conclude that

$$
0 \leqslant f_1(n) \leqslant 1 + O_{\epsilon,M,q}(\eta) \text{ for all } n \in \mathbb{Z}_N.
$$

Furthermore from (5.4) we have

$$
\mathbb{E}(f_1) = \mathbb{E}(f) \geqslant \delta.
$$

Applying Theorem 5.1 (possibly modifying f_1 by $O_{\epsilon,M,q}(\eta)$ first) we obtain

$$
(5.12) \qquad \mathbb{E}_{x,d\in\mathbb{Z}_N} f_1(x) f_1(x + d) f_1(x + 2d) \geqslant c(\delta) - O_{\epsilon,\delta,M,q}(\eta).
$$

Now we consider the other seven terms of the form (5.11), when at least one of the i, j, k is equal to 2. The treatment of the seven cases is essentially the same, so for simplicity we consider only $i = j = k = 2$. Observe that for all $a \in \mathbb{Z}_N$ we have

$$
(5.13) \qquad
\begin{aligned}
\widehat{f_1}(a) &= \mathbb{E}_{n\in\mathbb{Z}_N} \mathbb{E}_{m_1,m_2\in B} f(n + m_1 - m_2) e_N(an) \\
&= \mathbb{E}_{n\in\mathbb{Z}_N} \mathbb{E}_{m_1,m_2\in B} f(n + m_1 - m_2) e_N(a(n + m_1 - m_2) \times \\
&\qquad \times \ e_N(-am_1) e_N(am_2) \\
&= \widehat{f}(a)|\mathbb{E}_{m\in B} e(-am)|^2,
\end{aligned}
$$

and thus

$$(5.14) \qquad |\widehat{f_2}(a)| = |\widehat{f}(a)|(1 - |\mathbb{E}_{m \in B} e_N(-am)|^2) \leqslant |\widehat{f}(a)|.$$

This can be analysed in two different ways according as a does or does not lie in Ω. When $a \notin \Omega$ we have

$$|\widehat{f_2}(a)| \leqslant |\widehat{f}(a)| \leqslant \epsilon$$

by (5.8). On the other hand, when $a \in \Omega$, then $|e_N(-am) - 1| \leqslant \epsilon$ for all $m \in B$ by definition (5.9), and hence in this case we have

$$|\widehat{f_2}(a)| = |\widehat{f}(a)|(1 - |\mathbb{E}_{m \in B} e_N(-am)|^2) \leqslant 3\epsilon \mathbb{E}(\nu) \leqslant 3\epsilon(1 + \eta)$$

thanks to the $a = 0$ case of (5.5). Thus certainly

$$(5.15) \qquad\qquad\qquad \|\widehat{f_2}\|_\infty = O_\eta(\epsilon).$$

Now by orthogonality we obtain the well-known identity

$$\mathbb{E}_{n,d \in \mathbb{Z}_N} f_2(n) f_2(n + d) f_2(n + 2d) = \sum_{a \in \mathbb{Z}_N} \widehat{f_2}(a)^2 \widehat{f_2}(-2a).$$

Such an expression is often treated using an L^2–L^∞ inequality, but for us this is too expensive since we lack good L^2 control on ν (and hence on f_2). Instead, we shall use the L^q hypothesis (5.6). By Hölder's inequality we obtain

$$(5.16) \qquad |\mathbb{E}_{n,d \in \mathbb{Z}_N} f_2(n) f_2(n + d) f_2(n + 2d)| \leqslant \|\widehat{f_2}\|_q^q \|\widehat{f_2}\|_\infty^{3-q}.$$

From (5.6) and (5.14) we have

$$\|\widehat{f_2}\|_q \leqslant \|\widehat{f}\|_q \leqslant M.$$

Recalling (5.15) we therefore obtain, by (5.16), the bound

$$\mathbb{E}_{n,d \in \mathbb{Z}_N} f_2(n) f_2(n + d) f_2(n + 2d) = O(M^q \epsilon^{3-q}).$$

By choosing ϵ sufficiently small depending on δ, M, q, this can be made less than $\frac{1}{14} c(\delta)$. The same goes for the six other terms of the form (5.11) with $(i, j, k) \neq (1, 1, 1)$. Combining this with (5.12), the claim follows. $\qquad \square$

6. Arithmetic progressions of Chen primes

In this section we prove Theorem 1.2, which stated that there are infinitely many 3-term progressions of Chen primes. A Chen prime, recall, is a prime p such that $p + 2$ is a product of at most two primes, and Chen's famous theorem [6] is that there are infinitely many such primes. Chen actually proved a somewhat stronger result, in which the smallest prime factor of $p + 2$ can be bounded below by $p^{1/10}$. It is important for us to have this extra information (though any positive number would do in place of $1/10$). In Iwaniec's unpublished notes [22] one may find a proof of Chen's theorem which leads to the value $\gamma = 3/11$. For the purposes of this section we will

use the term Chen prime to refer to a p for which $p + 2$ is either prime or a product $p_1 p_2$, with $p_1, p_2 > p^{3/11}$. We quote, then, the following result from [22].[3]

Theorem 6.1. *Let N be a large integer. Then the number of Chen primes in the interval $(N/2, N]$ is at least $c_1 N / \log^2 N$, for some absolute constant $c_1 > 0$.*

For our argument we do not need to know an exact value for c_1. To have a chance of deducing Theorem 1.2 from Proposition 5.1, we shall need a suitably pseudorandom measure ν on the cyclic group $\mathbb{Z}_N$ which majorises the Chen primes on $[N/4, N/2]$ in a reasonably efficient way. Unfortunately, the Chen primes are so irregularly distributed in progressions $a \pmod q$ (q small) that there is no ν which would obey the pseudorandomness condition (5.5) for an η which is small enough to be useful. To get around this difficulty, we shall employ a device which we call the W-trick. Let $t \gg 1$ be a very large real number (independent of N) to be specified later, and write $W = W_t := \prod_{p \leqslant t} p$. Let $X_W \subseteq \mathbb{Z}_W$ denote the residue classes $b \in \mathbb{Z}_W$ such that b and $b + 2$ are both coprime to W; observe that

$$(6.1) \qquad |X_W| = W \prod_{3 \leqslant p \leqslant t} \left(1 - \frac{2}{p}\right) \ll |W| / \log^2 t.$$

Note that all but $O(W)$ of the Chen primes lie in one of the residue classes associated to X_W. From Theorem 6.1 we thus have

$$\sum_{b \in X_W} |\{N/4 \leqslant n \leqslant N/2 : Wn + b \text{ is a Chen prime}\}| \gg \frac{WN}{\log^2(WN)} - O(W).$$

If we assume N sufficiently large depending on W (and hence on t), the right-hand side is $\gg WN / \log^2 N$. From (6.1) and the pigeonhole principle we may therefore choose $b \in X_W$ such that

$$(6.2) \qquad |X| \gg \frac{N \log^2 t}{\log^2 N},$$

where X is the set

$$X := \{N/4 \leqslant n \leqslant N/2 : Wn + b \text{ is a Chen prime}\}.$$

Fix $b \in X_W$ as above, and consider the polynomial $F(n) := (Wn+b)(Wn+ b + 2)$. To compute $\mathfrak{S}_F$, we observe that $\gamma(p) = 1$ when $p \leqslant t$ and $\gamma(p) =$

---

[3]Theorem 6.1 is stronger than the result stated in [22] in one tiny way - Iwaniec shows that there are many Chen primes in $[0, N]$, not $(N/2, N]$. However it is completely clear that the proof can be adapted in a simple way to cover this case. Alternatively, a pigeonhole argument could be used to show that $(N/2, N]$ contains lots of Chen primes for many N, which would suffice for our purposes.

$1 - 2/p$ otherwise. Since the Euler product $\prod_p (1 - 2/p)/(1 - 1/p)^2$ is convergent, we see by the same computation used in (6.1) that

$$(6.3) \qquad \log^2 t \ll \mathfrak{S}_F \ll \log^2 t.$$

Now let $R := \lfloor N^{1/20} \rfloor$ and let $\beta_R : \mathbb{Z} \to \mathbb{R}^+$ be the enveloping sieve associated to F by Proposition 3.1. We restrict this function to the set $\{1, \dots, N\}$, which we identify with $\mathbb{Z}_N$, and let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be the resulting function.

Lemma 6.1. *For all $a \in \mathbb{Z}_N$ and $\epsilon > 0$, we have*

$$\hat{\nu}(a) = \delta_{a,0} + O_\epsilon(t^{\epsilon-1})$$

where $\delta_{a,0}$ is the Kronecker delta function.

Proof. From (3.3) we have

$$\hat{\nu}(a) = \mathbb{E}_{n \in \mathbb{Z}_N} \beta_R(n) e_N(an) = \sum_{q \leqslant R^2} \sum_{\substack{b \in \mathbb{Z}_q \\ (b,q)=1}} w(b/q) \mathbb{E}_{n \in \mathbb{Z}_N} e_N(an) e_q(-bn).$$

When $1 < q \leqslant t$ we have $\gamma(q) = 1$, and hence $w(b/q) = 0$ by Proposition 3.1(iv). By (3.4) we thus have
(6.4)

$$\hat{\nu}(a) = \mathbb{E}_{n \in \mathbb{Z}_N} e_N(an) + O_\epsilon(t^{\epsilon-1} \sum_{\substack{t \leqslant q \leqslant R^2}} \sum_{\substack{b \in \mathbb{Z}_q \\ (b,q)=1}} |\mathbb{E}_{n \in \mathbb{Z}_N} e_N(an) e_q(-bn)|).$$

Observe that $\mathbb{E}_{n \in \mathbb{Z}_N} e_N(an)$ is just $\delta_{a,0}$. From standard estimates on the Dirichlet kernel we have

$$(6.5) \qquad |\mathbb{E}_{n \in \mathbb{Z}_N} e_N(an) e_q(-bn)| \ll \min(1, \frac{1}{N\|\frac{a}{N} - \frac{b}{q}\|}).$$

All the fractions b/q, $q \leqslant R^2$, are separated from each other by at least $1/R^4 \leqslant 1/N$. Thus we have $1/N\|\frac{a}{N} - \frac{b}{q}\| \ll R^4/N$ with at most one exception, which leads in view of (6.4) and (6.5) to

$$\hat{\nu}(a) = \delta_{a,0} + O_\epsilon(t^{\epsilon-1}(1 + R^4 \frac{R^4}{N})).$$

The claim follows. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Theorem 1.2. We fix t to be an absolute constant to be chosen later, and take N to be a sufficiently large prime depending on t. We define the function $f : \mathbb{Z}_N \to \mathbb{R}^+$ by

$$f(n) := c \frac{\log^2 N}{\log^2 t} \mathbf{1}_X(n),$$

and verify the various conditions of Proposition 5.1 for f.

First of all observe from Proposition 3.1(i) and (6.3) that we have the majorization property (5.3) if $c > 0$ is chosen sufficiently small. Secondly, from (6.2) we have the condition (5.4) with some absolute constant $\delta > 0$ independent of N or t, as long as N is large enough. From Proposition 4.2 (specifically (4.7)) and Lemma 4.1 we obtain the bound $\|\widehat{f}\|_{5/2} \leqslant M$ for some $M > 0$ independent of N or t, if N is sufficiently large depending on t. This confirms condition (5.6) (with $q = 5/2$).

Finally, from Lemma 6.1 we obtain (5.5) with $\eta = O_\epsilon(t^{\epsilon-1})$ and $M_0 = 1 + O_\epsilon(t^{\epsilon-1})$.

Thus we may indeed apply Proposition 5.1, and doing so we obtain

$$\mathbb{E}_{x,d\in\mathbb{Z}_N} f(x)f(x+d)f(x+2d) \geqslant c - O_\epsilon(t^{\epsilon-1})$$

for some absolute constant $c > 0$. By choosing $\epsilon := 1/2$ (say) and t sufficiently large, we thus obtain

$$\mathbb{E}_{x,d\in\mathbb{Z}_N} f(x)f(x+d)f(x+2d) \geqslant c/2.$$

From the definition of f, this yields

$$|\{(x,d) \in \mathbb{Z}_N^2 : x, x+d, x+2d \in X\}| \gg_t N^2/\log^6 N.$$

The degenerate case $d = 0$ yields at most $O(N)$ pairs and can be discarded when N is large. Since X is contained in $[N/4, N/2]$ we see that $x, x + d, x + 2d$ is an arithmetic progression in $\mathbb{Z}$ and not just in $\mathbb{Z}_N$. Theorem 1.2 follows. $\qquad\square$

Remark. We have in fact shown that the number of triples (p_1, p_2, p_3) of Chen primes with $p_1 + p_3 = 2p_2$ and with each $p_i \leqslant N$ is $\gg N^2/\log^6 N$; it is possible to show from (1.8) that this bound is sharp up to an absolute multiplicative constant. It is clear that if one had a lower bound $\pi_2(N) \gg N/\log^2 N$ for the number of twin primes less than N (this, of course, is one of the most famous open conjectures in prime number theory) then a simple adaptation of our argument would produce infinitely many triples (p_1, p_2, p_3) of twin primes in arithmetic progression, and the number of such triples less than N would be $\gg N^2/\log^6 N$. Similar remarks hold for any other constellation of primes.

7. Appendix: Construction of the enveloping sieve

In this appendix we construct the enveloping sieve for Proposition 3.1 and verify its properties. Some of the material here is rather standard and will not be used directly in the rest of the paper, except via Proposition 3.1.

Let us recall the setup. We took a fixed polynomial F and defined the "locally sieved" sets

$$X_q = \{n \in \mathbb{Z}_q : (q, F(n)) = 1\}.$$

We begin by analyzing these locally sieved sets X_q. From (1.2) we have

(7.1) $$\gamma(q) = \mathbb{E}_{n \in \mathbb{Z}_q} \mathbf{1}_{(q,F(n))=1} = \mathbb{E}(\mathbf{1}_{X_q}) = \frac{1}{q}|X_q|.$$

From the Chinese remainder theorem we see that $|X_{qq'}| = |X_q||X_{q'}|$ whenever $(q, q') = 1$, and also that $|X_{p^m}| = p^{m-1}|X_p|$ whenever p^m is a prime power. From this and (7.1) we see that γ is a multiplicative function, and indeed

(7.2) $$\gamma(q) = \prod_{p|q} \gamma(p).$$

Observe that if $q|M$ then the projection $n \mapsto n(\mathrm{mod}\, q)$ will map X_M to X_q. Indeed, this map has uniform fibres:

Lemma 7.1. *If $q|M$, then*

$$\mathbb{E}_{n \in \mathbb{Z}_M, n \equiv m(\mathrm{mod}\, q)} \mathbf{1}_{X_M}(n) = \frac{\gamma(M)}{\gamma(q)} \mathbf{1}_{X_q}(m)$$

for all $m \in \mathbb{Z}_q$.

Proof. Let r be the largest factor of M which is coprime to q. We observe that if $n(\mathrm{mod}\, q) = m$, then

$$\mathbf{1}_{X_M}(n) = \mathbf{1}_{X_q}(m)\mathbf{1}_{X_r}(n(\mathrm{mod}\, r)).$$

From the Chinese remainder theorem we thus see that

$$\mathbb{E}_{n \in \mathbb{Z}_M, n \equiv m(\mathrm{mod}\, q)} \mathbf{1}_{X_M}(n) = \mathbf{1}_{X_q}(m)\mathbb{E}(\mathbf{1}_{X_r}) = \mathbf{1}_{X_q}(m)\gamma(r)$$

and the claim follows from (7.2). $\qquad\square$

We now analyze the Fourier behaviour of the X_q. For each $q \in \mathbb{N}$ and $a \in \mathbb{Z}$ we make the temporary definition

(7.3) $$t(a, q) := \mathbb{E}_{n \in X_q} e_q(an) = \frac{1}{q\gamma(q)} \sum_{n \in X_q} e_q(an).$$

Note that the denominator is non-zero by (1.3).

Lemma 7.2. *If $q \in \mathbb{Z}^+$ and $a \in \mathbb{Z}_q^*$, then $t(a, q) = t(ak, qk)$ for any $k \in \mathbb{Z}^+$.*

Proof. Let k' be the largest factor of k which is coprime to q. Observe that $n \in X_{qk}$ if and only if $n(\mathrm{mod}\, q) \in X_q$ and $n(\mathrm{mod}\, k') \in X_{k'}$. Since $e_q(an)$ depends only on $n(\mathrm{mod}\, q)$, we thus see from the Chinese remainder theorem that

$$\mathbb{E}_{n \in X_{qk}} e_q(an) = \mathbb{E}_{n \in X_q} e_q(an),$$

and the claim follows. □

As a consequence of this we may define the *normalised Fourier coefficients* $s(a/q)$ by

(7.4) $$s(a/q) := t(a, q).$$

We now record some useful identities and estimates for $s(a/q)$.

Lemma 7.3. *Let* $q, q' \in \mathbb{Z}^+$ *and* $a, a' \in \mathbb{Z}$. *Then the following statements hold.*

(i) $s(0) = 1$.
(ii) *s is periodic modulo 1. In particular when considering $s(a/q)$, we may think of a as an element of $\mathbb{Z}_q$.*
(iii) *If $(q, q') = 1$, then*
$$s(a/q + a'/q') = s(a/q)s(a'/q').$$

(iv) *If $a \in \mathbb{Z}_q^*$ and q is not square-free, $s(a/q) = 0$.*
(v) *If $a \in \mathbb{Z}_q^*$, $q > 1$, and $\gamma(q) = 1$, then $s(a/q) = 0$.*
(vi) *If $a \in \mathbb{Z}_q^*$, then $|s(a/q)| \leqslant \prod_{p|q}(1 - \gamma(p))/\gamma(p)$.*

Proof. The first two claims are immediate from (7.4). To confirm (iii), define $b(\mathrm{mod}\, qq')$ by the congruences $b \equiv aq'(\mathrm{mod}\, q)$ and $b \equiv a'q(\mathrm{mod}\, q')$, and r, r' by the relations $qr \equiv 1(\mathrm{mod}\, q')$, $q'r' \equiv 1(\mathrm{mod}\, q)$. Then it is easy to check that

(7.5) $$s(a/q)s(a'/q') = \frac{1}{|X_q||X_{q'}|} \sum_{n \in X_q} \sum_{n' \in X_{q'}} e_{qq'}\left(b(q'r'n + qrn')\right).$$

Observe that $q'r'n + qrn'$ is congruent to $n(\mathrm{mod}\, q)$ and to $n'(\mathrm{mod}\, q')$. Hence the right-hand side of (7.5) is indeed just

$$\frac{1}{|X_{qq'}|} \sum_{m \in X_{qq'}} e_{qq'}(bm) = s(b/qq') = s(a/q + a'/q').$$

To prove (iv), we write $q = p^2 k$ for some prime p and integer k, and observe that $X_{p^2 k}$ is periodic mod pk. Since $\frac{a}{q}$ is not an integer divided by pk, we see that the expression in (7.4) necessarily vanishes. The fifth claim (v) follows by inspection. Finally, we prove (vi). By (iv) we may assume q is square-free. By (iii) we may assume q is prime. But then (vi) follows by writing

$$\sum_{n \in X_q} e_q(an) = \sum_{n \in \mathbb{Z}_q} e_q(an) - \sum_{n \in \mathbb{Z}_q \backslash X_q} e_q(an),$$

and observing that the former sum vanishes since $a \in \mathbb{Z}_q^*$, while the latter sum is at most $|\mathbb{Z}_q \backslash X_q| = q(1 - \gamma(q))$ by (7.1). □

Now we can write $\mathbf{1}_{X_M}$ as a Fourier series, with coefficients determined by the γ and s functions.

Lemma 7.4. *For any M and any $n \in \mathbb{Z}_M$, we have*

$$(7.6) \qquad \mathbf{1}_{X_M}(n) = \gamma(M) \sum_{q|M} \sum_{a \in \mathbb{Z}_q^*} s(a/q) e_q(-an).$$

Proof. The right-hand side can be rewritten as $\gamma(M) \sum_{b \in \mathbb{Z}_M} s(b/M) e_M(-bn)$. But from (7.3) and (7.4) we have $\gamma(M) s(b/M) = \mathbb{E}_{n \in \mathbb{Z}_M} \mathbf{1}_{X_M}(n) e_M(bn)$. The claim then follows from the Fourier inversion formula. $\qquad \square$

Motivated by the identity (7.6), let us define the function[4] $\alpha_R : \mathbb{Z} \to \mathbb{R}$ by

$$(7.7) \qquad \alpha_R(n) := \sum_{q \leqslant R} \sum_{a \in \mathbb{Z}_q^*} s(a/q) e_q(-an).$$

Thus α_R is some sort of truncated version of $\mathbf{1}_{X_{R!}}/\gamma(R!)$, which is normalized to have average value 1 (as can be seen by considering the $q = 1$ term of the above sum, which is the only term which does not oscillate in n).

Lemma 7.5. *Let h denote the multiplicative function*

$$(7.8) \qquad h(q) := \mu(q)^2 \prod_{p|q} \frac{1 - \gamma(p)}{\gamma(p)}.$$

Then for all $n \in X_{R!}$, we have $\alpha_R(n) = G(R)$, where $G(R) = G(F, R)$ is the quantity

$$(7.9) \qquad G(R) := \sum_{q \leqslant R} h(q).$$

Remark. Note in particular that $G(R)$ is positive (the $q = 1$ summand is equal to 1).

Proof. Fix $n \in X_{R!}$. Recall the definition (7.7) of α_R. In the first instance we will work with the inner sum $\sum_{a \in \mathbb{Z}_q^*} s(a/q) e_q(-an)$ occurring in that definition. We may restrict q to be square-free since only such q contribute

---

[4]While at first glance this expression appears to be complex-valued, it is in fact real, as one can see from conjugation symmetry, or from more explicitly real formulae for α_R below such as (7.10).

to the sum (7.7) in view of Lemma 7.3 (iv). We have

$$
\begin{aligned}
\sum_{a \in \mathbb{Z}_q^*} s(a/q) e_q(-an) &= \sum_{a \in \mathbb{Z}_q^*} \mathbb{E}_{m \in X_q} e_q(am) e_q(-an) \\
&= \mathbb{E}_{m \in X_q} \sum_{a \in \mathbb{Z}_q} e_q(a(m-n)) \mathbf{1}_{(a,q)=1} \\
&= \mathbb{E}_{m \in X_q} \sum_{a \in \mathbb{Z}_q} e_q(a(m-n)) \sum_{d \mid (a,q)} \mu(d) \\
&= \sum_{d \mid q} \mu(d) \mathbb{E}_{m \in X_q} \sum_{a \in \mathbb{Z}_q : d \mid a} e_q(a(m-n)) \\
&= \sum_{d \mid q} \mu(d) \mathbb{E}_{m \in X_q} \frac{q}{d} \mathbf{1}_{m \equiv n \,(\mathrm{mod}\, q/d)} \\
&= \sum_{d \mid q} \frac{\mu(d) \mathbb{E}_{m \in \mathbb{Z}_q, m \equiv n \,(\mathrm{mod}\, q/d)} \mathbf{1}_{X_q}(m)}{\mathbb{E}_{m \in X_q} \mathbf{1}_{X_q}(m)}.
\end{aligned}
$$

By Lemma 7.1, we have

$$
\frac{\mathbb{E}_{m \in \mathbb{Z}_q, m \equiv n \,(\mathrm{mod}\, q/d)} \mathbf{1}_{X_q}(m)}{\mathbb{E}_{m \in \mathbb{Z}_q} \mathbf{1}_{X_q}(m)} = \frac{1}{\gamma(q/d)} \mathbf{1}_{X_{q/d}}(n).
$$

Now we substitute into (7.7), the definition of α_R. Since q is being supposed square-free, d and q/d are coprime. Hence

$$
(7.10) \qquad \alpha_R(n) = \sum_{q \leqslant R} \mu^2(q) \sum_{d \mid q : (\frac{q}{d}, F(n)) = 1} \frac{\mu(d)\gamma(d)}{\gamma(q)}.
$$

Since $n \in X_R$, we have $(\frac{q}{d}, F(n)) = 1$ for all $q \leqslant R$ and $d \mid q$. Furthermore from the multiplicativity of γ we have

$$
\sum_{d \mid q} \frac{\mu(d)\gamma(d)}{\gamma(q)} = \prod_{p \mid q} \frac{1 - \gamma(p)}{\gamma(p)} = h(q),
$$

and the claim follows. □

The function α_R is not quite a candidate for an enveloping sieve, because it can be negative. To resolve this problem, we shall simply square (and renormalize) α_R, defining

$$
(7.11) \qquad \beta_R(n) := \frac{1}{G(R)} |\alpha_R(n)|^2.
$$

Clearly β_R is non-negative, has $R!$ as a period, and equals $G(R)$ on $X_{R!}$. We now give a Fourier representation for β_R similar to (7.7), but with slightly larger Fourier coefficients.

Proposition 7.1. [28] *We have the representation*

$$\beta_R(n) = \sum_{q \leqslant R^2} \sum_{a \in \mathbb{Z}_q^*} w(a/q) e_q(-an),$$

where the coefficients $w(a/q)$ obey the bounds

$$(7.12) \qquad\qquad |w(a/q)| \leqslant 3^{\varpi(q)} |s(a/q)|$$

(here $\varpi(q)$ denotes the number of prime factors of q). Also we have $w(0) = 1$.

Remark. More precise asymptotics for $w(a/q)$ are available, see [28] (and see [27] for an even more precise statement in the case $F(n) = n$). We will not need these refinements here, however.

Proof. We follow the arguments of Ramaré and Ruzsa [28]. Our starting point is the formula (7.10), which was derived for all n (not just those in $X_{R!}$). Writing $q = dr$, we thus have[5],

$$(7.13) \qquad\qquad \alpha_R(n) = \sum_{d,r:dr \leqslant R} \frac{\mu(d)}{\gamma(r)} \mathbf{1}_{(r,F(n))=1}.$$

Inserting (7.13) into (7.11), we obtain

$$\beta_R(n) = \frac{1}{G(R)} \sum_{\substack{d_1,r_1,d_2,r_2 \\ d_1 r_1, d_2 r_2 \leqslant R}} \frac{\mu(d_1)\mu(d_2)}{\gamma(r_1)\gamma(r_2)} \mathbf{1}_{([r_1,r_2],F(n))=1},$$

where $[r_1, r_2]$ denotes the least common multiple of r and r'. Also note that α_R, and hence β_R, is periodic with period $R!$. By Fourier inversion, we then have

$$\beta_R(n) = \sum_{q|R!} \sum_{a \in \mathbb{Z}_q^*} w(a/q) e_q(-an)$$

where

$$w(a/q) := \frac{1}{G(R)} \mathbb{E}_{n \in \mathbb{Z}_{R!}} \sum_{\substack{d_1,r_1,d_2,r_2 \\ d_1 r_1, d_2 r_2 \leqslant R}} \frac{\mu(d_1)\mu(d_2)}{\gamma(r_1)\gamma(r_2)} \mathbf{1}_{([r_1,r_2],F(n))=1} e_q(an).$$

Fix a, q. Observe that the quantity $\mathbf{1}_{([r_1,r_2],F(n))=1}$, as a function of n, is periodic with period $[r_1, r_2]$. Thus $\mathbb{E}_{n \in \mathbb{Z}_{R!}} \mathbf{1}_{([r_1,r_2],F(n))=1} e_q(an)$ vanishes unless q divides $[r_1, r_2]$, in which case the expression is equal to $s(a/q)\gamma([r_1,r_2])$

---

[5]Note that the contribution of the q which are not square free can be added in, since it is easy to show that $\sum_{dr=q} \frac{\mu(d)}{\gamma(r)} \mathbf{1}_{(r,F(n))=1} = 0$ for all such q. Indeed, if $p^2|q$, then the contributions of the cases $p|d, p|r$ and $p \nmid d, p^2|r$ to this sum will cancel each other out.

by (7.4). In particular, since $[r_1, r_2] \leqslant R^2$, we know that $w(a/q)$ vanishes when $q > R^2$, and in the remaining cases $q \leqslant R^2$ we have

$$w(a/q) = \frac{s(a/q)}{G(R)} \sum_{\substack{d_1, r_1, d_2, r_2 \\ d_1 r_1, d_2 r_2 \leqslant R \\ q \mid [r_1, r_2]}} \frac{\mu(d_1)\mu(d_2)}{\gamma(r_1)\gamma(r_2)} \gamma([r_1, r_2]).$$

From (7.2) and (7.8) we have

$$\frac{\gamma([r_1, r_2])}{\gamma(r_1)\gamma(r_2)} = \prod_{p \mid (r_1, r_2)} \frac{1}{\gamma(p)} = \prod_{p \mid (r_1, r_2)} (1 + h(p)) = \sum_{t \mid (r_1, r_2)} h(t),$$

and thus

(7.14) $$w(a/q) = \frac{s(a/q)}{G(R)} \sum_{t \leqslant R} h(t) \sum_{\substack{d_1, r_1, d_2, r_2 \\ d_1 r_1, d_2 r_2 \leqslant R \\ q \mid [r_1, r_2] \\ t \mid (r_1, r_2)}} \mu(d)\mu(d').$$

Thus to prove (7.12), it suffices by (7.9) to show that

$$\left| \sum_{\substack{d_1, r_1, d_2, r_2 \\ d_1 r_1, d_2 r_2 \leqslant R \\ q \mid [r_1, r_2] \\ t \mid (r_1, r_2)}} \mu(d_1)\mu(d_2) \right| \leqslant 3^{\varpi(q)}$$

for all $t \leqslant R$. While this is in fact true for all q, it in fact suffices in light of Proposition 7.3 (iv) to verify it for square-free q, in which case the formulae are slightly simpler.

Fix t. We take advantage of the constraint $t \mid (r_1, r_2)$ to write $r_j = t\tilde{r}_j$, $l_j = d_j \tilde{r}_j$, and $\tilde{q} := q/(q, t)$, and observe that the constraint $q \mid [r_1, r_2]$ is equivalent to $\tilde{q} \mid [\tilde{r}_1, \tilde{r}_2]$. Also, $\tilde{q}$ is square-free. Thus we reduce to proving that

(7.15) $$\left| \sum_{l_1, l_2 \leqslant R/t} A(l_1, l_2) \right| \leqslant 3^{\varpi(q)}$$

where

$$A(l_1, l_2) := \sum_{\substack{\tilde{r}_1 \mid l_1, \tilde{r}_2 \mid l_2 \\ \tilde{q} \mid [\tilde{r}_1, \tilde{r}_2]}} \mu(l_1/\tilde{r}_1)\mu(l_2/\tilde{r}_2).$$

This expression can be worked out quite explicitly:

Lemma 7.6. *Let $\tilde{q}$ be square-free. Then the quantity $A(l_1, l_2)$ vanishes unless $l_1, l_2 \mid \tilde{q}$ and $\tilde{q} \mid l_1 l_2$, in which case $|A(l_1, l_2)| = 1$.*

Proof.   It suffices to prove this claim in the "local" case when $l_1, l_2, \tilde{q}$ are all powers of a single prime p, since the general case then follows by splitting $l_1, l_2, \tilde{q}$ into prime factors and exploiting the multiplicativity of the summation in $A(l_1, l_2)$. If $\tilde{q} = 1$ then the constraint $\tilde{q}|[\tilde{r}_1, \tilde{r}_2]$ is vacuous and the claim follows from Möbius inversion. Since $\tilde{q}$ is square-free, the only remaining case is when $\tilde{q} = p$. In this case, $\tilde{q}|[\tilde{r}_1, \tilde{r}_2]$ if and only if not both of $\tilde{r}_1, \tilde{r}_2$ equal one, and so

$$A(l_1, l_2) = \sum_{\tilde{r}_1|l_1, \tilde{r}_2|l_2} \mu(l_1/\tilde{r}_1)\mu(l_2/\tilde{r}_2) - \mu(l_1)\mu(l_2).$$

Applying Möbius inversion (or direct computation) we thus see that the expression $A(l_1, l_2)$ equals -1 when $(l_1, l_2) = (p, p)$, equals $+1$ when $(l_1, l_2) = (p, 1), (1, p)$, and vanishes otherwise, and the claim follows.   □

From this lemma we see that there are at most $3^{\varpi(\tilde{q})} \leqslant 3^{\varpi(q)}$ pairs (l_1, l_2) for which $A(l_1, l_2)$ is non-zero, and in each of these cases we have $|A(l_1, l_2)| = 1$. The claim (7.15) follows. Using (7.14), this proves (7.12).

In the case $a = 0$ and $q = 1$, a closer inspection of the above argument shows that the quantity in absolute values in (7.15) is in fact exactly equal to 1. In view of (7.14), Lemma 7.3 (i) and (7.9), this gives $w(0) = 1$ as desired. This concludes the proof of Proposition 7.1.   □

Finally, we rewrite α_R (and hence β_R) in the more familiar notation of the Selberg sieve. This will allow us to conclude the crude estimate for β_R demanded in Proposition 3.1 (ii).

Lemma 7.7. *We have*

$$\alpha_R(n) = G(R) \sum_{\substack{d \leqslant R \\ d|F(n)}} \lambda_d,$$

where λ_d is defined for squarefree d by

(7.16) $$\lambda_d := \frac{\mu(d)G_d(R/d)}{\gamma(d)G(R)},$$

and the quantity G_d is defined by

$$G_d(x) := \sum_{\substack{q \leqslant x \\ (q,d)=1}} h(q).$$

Remark. The notation here is reasonably consistent with what is normal in sieve theory, as may be found for example in [19]. For some further discussion of the weights λ_d in the special case $F(n) = n$, see Appendix 8 below.

Proof. We start with (7.13) and use Möbius inversion to write

$$1_{(r,F(n))=1} = \sum_{m|r, m|F(n)} \mu(m),$$

and hence

$$\alpha_R(n) = \sum_{m \leqslant R : m|F(n)} \mu(m) \sum_{s,r : sr \leqslant R, m|r} \frac{\mu(s)}{\gamma(r)}.$$

Writing $r =: mr'$ and $q := sr'$, we thus have

$$\alpha_R(n) = \sum_{m \leqslant R : m|F(n)} \mu(m) \sum_{q \leqslant R/m} \sum_{r'|q} \frac{\mu(q/r')}{\gamma(mr')}.$$

From (7.2) a calculation gives

$$\sum_{r'|q} \frac{\mu(q/r')}{\gamma(mr')} = \frac{1_{(q,m)=1}}{\gamma(m)} \prod_{p|q} \mu^2(q)\left(\frac{1}{\gamma(q)} - 1\right) = \frac{1_{(q,m)=1}}{\gamma(m)} h(q),$$

and the claim follows. $\qquad \square$

Lemma 7.8. *Suppose that d is squarefree. Then $|\lambda_d| \leqslant 1$.*

Proof. This is a well-known fact in sieve theory. The proof (which may be found in [19, p. 100]) consists in observing that

$$G(R) = \sum_{l|d} h(l)G_d(R/l) \geqslant \left(\sum_{l|d} h(l)\right)G_d(R/d) = \frac{G_d(R/d)}{\gamma(d)},$$

which implies the result immediately. $\qquad \square$

Proof of Proposition 3.1. To prove (i), it suffices in view of Lemma 7.5 to show that

$$G(R) \gg_k \frac{\log^k R}{\mathfrak{S}_F}.$$

In [19, Lemma 4.1], the bound

$$G(R) \gg_k \prod_{p < R} \frac{1}{\gamma(p)}$$

is obtained. The claim (i) now follows from (1.7) and the classical bound $\prod_{k < p \leqslant R}(1 - k/p) \ll_k (\log R)^{-k}$.

Let us now consider (ii), that is the statement $\beta_R(n) \ll_\epsilon N^\epsilon$. By Lemmas 7.7 and 7.8 we have the bound $\beta(n) \leqslant G(R)d(F(n))^2$. Recall that $1 - \gamma(p) \leqslant k/p$ for all p (cf. (1.4)), together with the non-degeneracy assumption (1.3),

which implies that $\gamma(p) \geqslant 1/p$. For any $\epsilon > 0$, we may therefore infer the bound
(7.17)
$$h(q) = \mu(q)^2 \prod_{p|q} \frac{1 - \gamma(p)}{\gamma(p)} \leqslant k^k \mu(q)^2 \prod_{\substack{p|q \\ p>k}} \frac{k/p}{1 - k/p} \ll_k k^{\varpi(q)} \log^k q/q \ll_{k,\epsilon} q^{\epsilon-1}$$

for any $\epsilon > 0$. Here, we have invoked well-known inequalities for the divisor function (recall that $|a_i|, |b_i| \leqslant N$, whence $|F(n)| \ll_k N^{2k}$). Using (7.17), and recalling (7.9), we infer that

$$G(R) \ll_{k,\epsilon} R^{\epsilon/2} \leqslant N^{\epsilon/2},$$

concluding the proof of (ii). We remark that more precise bounds on $G(R)$ may be found in [19].

To prove (iii) and (iv), it suffices by Lemma 7.3 (vi) and Proposition 7.1 to show that

$$3^{\varpi(q)} \prod_{p|q} (1 - \gamma(p))/\gamma(p) \ll_\epsilon q^{\epsilon-1}$$

for all square-free q. But the left-hand side is just $3^{\varpi(q)} h(q)$, whence by (7.17) and the classical bound $\varpi(q) \ll \log q / \log \log q$ we are done. $\qquad\square$

8. Appendix: A comparison of two enveloping sieves.

Let us begin with a few further remarks concerning what we mean by an enveloping sieve. The idea of considering the sieve β_R not simply as a tool for obtaining upper bounds but as an interesting function in its own right is one we got by reading papers of Ramaré [27] and Ramaré-Ruzsa [28]. Ramaré generously acknowledges that the essential idea goes back to Hooley [20]. It is in those papers (and in fact so far as we know *only* in those papers) that one finds the term *enveloping sieve*.

The name, though perhaps nonstandard, seems to us to be appropriate for describing a situation where one uses a single majorant for the primes to deal simultaneously with a whole family of problems which might normally, by themselves, be dealt with by a family of different sieves.

In Proposition 3.1 take $F(n) = n$, so that $G(R)^{-1}\beta_R$, as defined in (7.11), is a majorant for the primes themselves (or at least the primes $> R$). The expansion of β_R as a Fourier series in Proposition 3.1 can be thought of as a way of using a sieve for the primes to tell us about the behaviour of the primes restricted to arithmetic progressions $a(\mathrm{mod}\, q)$. This might normally be handled by a separate procedure for each (a, q).

The authors [18] used a somewhat different majorant for the primes, which we call β'_R below (a closely related object was called ν in [18]). We were concerned with the estimation of correlations

$$(8.1) \qquad \mathbb{E}_{n \leqslant N} \beta'_R(n + h_1) \ldots \beta'_R(n + h_m).$$

Here, then, a single sieve β'_R is being used to deal with configurations $(n + h_1, \ldots, n + h_m)$ which might more normally be treated using a Selberg sieve adapted to the polynomial $F(n) = (n + h_1) \ldots (n + h_m)$. This, we think, should also qualify as a use of an enveloping sieve.

Let us return to the function $\beta_R(n)$. By Lemma 7.7 it is given by the formula

$$\beta_R(n) = G(R) (\sum_{\substack{d|n \\ d \leqslant R}} \lambda_d^{\mathrm{SEL}})^2.$$

$G(R)$ is as defined in (7.9). In this particular case we have from (7.8) that $h(m) = 1/\phi(m)$ and so

$$G(R) = \sum_{m \leqslant R} \frac{\mu^2(m)}{\phi(m)}.$$

The weights λ_d^{SEL} (SEL stands for Selberg) are defined by (7.16); when $F(n) = n$ we have $\gamma(d) = \phi(d)/d$, and thus

$$(8.2) \qquad \lambda_d^{\mathrm{SEL}} := \frac{\mu(d) G_d(R/d)}{\gamma(d) G(R)} = \frac{d\mu(d)}{\phi(d) G(R)} \sum_{\substack{q \leqslant R/d \\ (q,d)=1}} 1/\phi(q).$$

Why did we consider β_R? For us, the motivating factor was that the Fourier transform of β_R could be described quite accurately, as was done for example in Proposition 7.1. There are many other uses for β_R, one of the most famous being to the so-called Brun-Titchmarsh problem of proving an upper bound for the number of primes in an interval $[x, x + y)$. The function β_R and the associated weights λ_d^{SEL} were first considered by Selberg and constitute the simplest instance of Selberg's *upper bound sieve* (also known as the Λ^2-sieve).

As is well-known, the weights λ_d^{SEL} arise from a certain natural quadratic optimization problem. Indeed for *any* real weights λ_d such that $\lambda_1 = 1$ the function

$$\psi(n) := (\sum_{\substack{d|n \\ d \leqslant R}} \lambda_d)^2$$

forms a pointwise majorant for the primes $> R$. For a given integer N we might choose R to be a small power of N and then attempt to choose the λ_d so that $\sum_{n \leqslant N} \psi(n)$ is as small as possible. After changing the order of summation one finds that

$$(8.3) \qquad \sum_{n \leqslant N} \psi(n) = N \sum_{d_1, d_2 \leqslant R} \frac{\lambda_{d_1} \lambda_{d_2}}{[d_1, d_2]} + O(R^2 \sup_{d \leqslant R} \lambda_d^2).$$

In practice the error will be small, and we become interested in minimising the quadratic form

$$Q(\lambda) := \sum_{d_1, d_2 \leqslant R} \frac{\lambda_{d_1} \lambda_{d_2}}{[d_1, d_2]} = \sum_{\delta \leqslant R} \phi(\delta) (\sum_{\delta | d} \frac{\lambda_d}{d})^2$$

subject to $\lambda_1 = 1$. $Q(\lambda)$ may be written as $\sum_{\delta \leqslant R} \phi(\delta) u_\delta^2$ where

$$(8.4) \qquad u_\delta := \sum_{\delta | d} \lambda_d / d;$$

in these new coordinates the condition $\lambda_1 = 1$ translates to

$$\sum_{\delta \leqslant R} \mu(\delta) u_\delta = 1.$$

Observe that under these conditions we have

$$(8.5) \qquad Q(\lambda) = \sum_{\delta \leqslant R} \phi(\delta) \left(u_\delta - \frac{\mu(\delta)}{G(R)\phi(\delta)} \right)^2 + G(R)^{-1},$$

which makes it completely clear that the minimum of $Q(\lambda)$ is $1/G(R)$ and that this occurs when $u_\delta = \mu(\delta)/\phi(\delta)G(R)$. This corresponds to the choice $\lambda_d = \lambda_d^{\mathrm{SEL}}$.

Now it is well-known that $G(R) \sim \log R$. In fact one has the asymptotic

$$(8.6) \qquad G(R) = \log R + \gamma + \sum_p \frac{\log p}{p(p-1)} + o(1).$$

This may be found in [27], for example; Montgomery [25] traces the result back at least as far as Ward [37].

We remark that all this together with Lemma 7.8 (which, in our new notation, tells us that $|\lambda_d^{\mathrm{SEL}}| \leqslant 1$) allows us to recover the fact that if $R = N^{1/2-\epsilon}$ then

$$\sum_{n \leqslant N} \beta_R(n) = N(1 + o_\epsilon(1)).$$

This also follows quickly from Proposition 3.1 (cf. Lemma 4.1).

As remarked earlier, we had occasion in our paper [18, Chapter 8] to consider a somewhat different majorant for the primes $> R$. To enable a comparison with β_R, we write this majorant in the form

$$(8.7) \qquad \beta'_R(n) = G(R)(\sum_{\substack{d|n \\ d \leqslant R}} \lambda_d^{\mathrm{GY}})^2$$

where

$$\lambda_d^{\mathrm{GY}} = \frac{\mu(d)\log(R/d)}{\log R}.$$

Objects of this type were extensively analysed by Goldston and Yıldırım [11, 12, 13], and in particular they saw how to asymptotically evaluate certain *correlations* of the form (8.1) provided $R \leqslant N^{c_m}$. This was a crucial ingredient in our work in [18], but we should remark that other aspects of the function β'_R were investigated much earlier, and indeed β'_R was known to Selberg. It has also featured in works of Friedlander-Goldston [8], Goldston [10] and Hooley [21], among others. The consideration of the weights λ^{GY} and the associated majorant β'_R may be motivated by looking at an asymptotic for the weights λ_d^{SEL}. In [27, Lemma 3.4], for example, one finds the result

$$(8.8) \qquad G_d(R/d) = \log(R/d) + \gamma + \sum_{p \geqslant 2} \frac{\log p}{p(p-1)} + \sum_{p|d} \frac{\log p}{p},$$

which leads in view of the definition (8.2) and (8.6) to the asymptotic

$$\lambda_d^{\mathrm{SEL}} \sim \frac{\mu(d)\log(R/d)}{\log R} = \lambda_d^{\mathrm{GY}}.$$

At the very least, this suggests looking at β'_R as defined in (8.7) above. The oscillatory effect of the Möbius function means that there is no *a priori* guarantee that β'_R is a close approximation to β_R. It turns out however that β_R and β'_R are, for many purposes, rather similar.

One way to see this is to inspect the quadratic form $Q(\lambda^{\mathrm{GY}})$. The asymptotic

$$Q(\lambda^{\mathrm{GY}}) = \frac{1}{\log R} + O(\frac{1}{\log^2 R})$$

follows from a result of Graham [15] and can easily be proved using Goldston and Yıldırım's method (as exposited, for example, in [18, Ch. 9]). Comparing with (8.5) and using the asymptotic (8.6) one sees that

$$\sum_{\delta \leqslant R} \phi(\delta)(u_\delta^{\mathrm{GY}} - u_\delta^{\mathrm{SEL}})^2 = O(\frac{1}{\log^2 R}).$$

Here, of course, u^{GY} and u^{SEL} are related to λ^{GY} and λ^{SEL} respectively by the transformation (8.4). If $R \ll N^{1/2-\epsilon}$ then this translates via (8.3) to a bound

$$\sum_{n \leqslant N} \Big(\sum_{\substack{d|n \\ d \leqslant R}} \lambda_d^{\text{GY}} - \lambda_d^{\text{SEL}} \Big)^2 = O(\frac{N}{\log^2 R}),$$

which implies that

$$\sum_{n \leqslant N} \Big(\beta_R'^{1/2}(n) - \beta_R^{1/2}(n) \Big)^2 = O(\frac{N}{\log R}).$$

Using the Cauchy-Schwarz inequality we now get

$$
\begin{aligned}
\|\beta_R' - \beta_R\|_1 \quad &:= \quad \mathbb{E}_{1 \leqslant n \leqslant N} |\beta_R'(n) - \beta_R(n)| \\
&\leqslant \quad \big(\mathbb{E}_{1 \leqslant n \leqslant N} (\beta_R'^{1/2}(n) - \beta_R^{1/2}(n))^2 \big)^{1/2} \times \\
&\qquad \times \big(\mathbb{E}_{1 \leqslant n \leqslant N} (\beta_R'^{1/2}(n) + \beta_R^{1/2}(n))^2 \big)^{1/2} \\
&= \quad O(\frac{1}{\log R})^{1/2} \cdot (\|\beta_R\|_1 + \|\beta_R'\|_1)^{1/2} \\
&= \quad O(\frac{1}{\sqrt{\log R}}).
\end{aligned}
$$

On the other hand, from Lemma 4.1 we have $\|\beta_R\|_1 = O(1)$. From the triangle inequality we thus obtain

Lemma 8.1. *Suppose that $R \ll N^{1/2-\epsilon}$. Then β_R and β_R' are close in that they satisfy the L^1-estimate $\|\beta_R' - \beta_R\|_1 = O((\log R)^{-1/2})$.* $\qquad\square$

Such an estimate is not, in itself, enough to allow us to deduce enough information about the Fourier expansion of β_R' to prove a restriction theorem analogous to that obtained in §4 for β_R. Such Fourier information could, we believe, be obtained by an analytic method similar to that in [13], but there seems little use for such a result. Going in the other direction, one might ask whether it is possible to obtain combinatorial information concerning correlations of the form (8.1) for β_R. In view of (8.8) it looks as though any such effort might require one to address the function

$$H_R(n) := \sum_{\substack{d|n \\ d \leqslant R}} \mu(d),$$

concerning oneself in particular with the $2m^{\text{th}}$ moment of this function. When $m = 1$ this was investigated by Dress, Iwaniec and Tenenbaum [7] and when $m = 2$ by Motohashi [26]. At this point there are already some rather thorny issues involved, and it does not seem worth the effort of pursuing the matter simply to obtain correlation estimates for β_R. The

function H_R is of some interest in its own right, however, and had an auxilliary role for instance in the ground-breaking work [9].

Let us conclude with the following remark. As far as enveloping sieves for the primes are concerned, it seems that both β_R and β'_R have their own role to play. β_R is useful if one wants to do harmonic analysis, whereas β'_R is far more appropriate if one wishes to do combinatorics. The two majorants are, however, rather closely related.

References

[1] G.F. Bachelis, *On the upper and lower majorant properties in $L^p(G)$,* Quart. J. Math. (Oxford) (2) **24** (1973), 119–128.

[2] A. Balog, *The Hardy-Littlewood k-tuple conjecture on average,* Analytic Number Theory (eds. B. Brendt, H.G. Diamond, H. Halberstam and A. Hildebrand), Birkhäuser, 1990, 47–75.

[3] J. Bourgain, *On $\Lambda(p)$-subsets of squares,* Israel J. Math. **67** (1989), 291–311.

[4] ———, *Fourier restriction phenomena for certain lattice subsets and application to non-linear evolution equations I. Schrödinger equations,* GAFA **3** (1993), 107–156.

[5] ———, *On triples in arithmetic progression,* GAFA **9** (1999), no. 5, 968–984.

[6] J.-R. Chen, *On the representation of a large even integer as the sum of a prime and a product of at most two primes,* Sci. Sinica **16** (1973), 157–176.

[7] F. Dress, H. Iwaniec and G. Tenenbaum, *Sur une somme liée à la fonction Möbius,* J. reine angew. Math. **340** (1983), 53–58.

[8] J.B. Friedlander and D.A. Goldston, *Variance of distribution of primes in residue classes,* Quart. J. Math. (Oxford) (2) **47** (1996), 313–336.

[9] J.B. Friedlander and H. Iwaniec, *The polynomial $X^2 + Y^4$ captures its primes,* Ann. Math. **148** (1998), no. 3, 965–1040.

[10] D.A. Goldston, *A lower bound for the second moment of primes in short intervals,* Exposition Math. **13** (1995), no. 4, 366–376.

[11] D.A. Goldston and C.Y. Yıldırım, *Higher correlations of divisor sums related to primes, I: Triple correlations,* Integers **3** (2003) A5, 66pp.

[12] D.A. Goldston and C.Y. Yıldırım, *Higher correlations of divisor sums related to primes, III: k-correlations,* preprint. Available at:
http://www.arxiv.org/pdf/math.NT/0209102.

[13] D.A. Goldston and C.Y. Yıldırım, *Small gaps between primes I,* preprint.

[14] W.T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four,* GAFA **8** (1998), 529–551.

[15] S.W. Graham, *An asymptotic estimate related to Selberg's sieve,* J. Num. Th. **10** (1978), 83–94.

[16] B.J. Green, *Roth's theorem in the primes,* to appear in Annals of Math. **161** (2005), no. 3.

[17] B.J. Green and I.Z. Ruzsa, *On the Hardy-Littlewood majorant problem,* Math. Proc. Camb. Phil. Soc. **137** (2004), no. 3, 511–517.

[18] B.J. Green and T.C. Tao, *The primes contain arbitrarily long arithmetic progressions,* preprint.

[19] H. Halberstam and H.E. Richert, *Sieve Methods,* London Math. Soc. Monographs **4**, Academic Press 1974.

[20] C. Hooley, *Applications of sieve methods in number theory,* Cambridge Tracts in Math. **70**, CUP 1976.

[21] C. Hooley, *On the Barban-Davenport-Halberstam theorem, XII,* Number Theory in Progress, Vol. 2 (Zakopan–Košcielisko 1997), 893–910, de Gruyter, Berlin 1999.

[22] H. Iwaniec, *Sieve methods,* Graduate course, Rutgers 1996.

[23] G. Mockenhaupt and W. Schlag, manuscript.

[24] H.L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS Regional Conference Series in Mathematics **84**, AMS, Providence, RI, 1994.

[25] H.L. Montgomery, *Selberg's work on the zeta-function,* in Number Theory, Trace Formulas and Discrete Groups (a Symposium in Honor of Atle Selberg, Oslo, Norway, July 14–21, 1987).

[26] Y. Motohashi, *A multiple sum involving the Möbius function,* Publ. Inst. Math. (Beograd) (N.S.) **76**(90) (2004), 31–39. Also available at
`http://www.arxiv.org/pdf/math.NT/0310064`.

[27] O. Ramaré, *On Snirel'man's constant,* Ann. Scu. Norm. Pisa **22** (1995), 645–706.

[28] O. Ramaré and I.Z. Ruzsa, *Additive properties of dense subsets of sifted sequences,* J. Th. Nombres de Bordeaux **13** (2001), 559–581.

[29] K.F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.

[30] I.Z. Ruzsa, *On an additive property of squares and primes,* Acta Arithmetica **49** (1988), 281–289.

[31] A. Selberg *On an elementary method in the theory of primes,* Kong. Norske Vid. Selsk. Forh. B **19** (18), 64–67.

[32] T.C. Tao, *Recent progress on the restriction conjecture,* in Fourier analysis and convexity (Milan, 2001), 217–243, Appl. Numer. Harmon. Anal., Birkhäuser Boston, Boston, MA, 2004.

[33] D.I. Tolev, *Arithmetic progressions of prime-almost-prime twins,* Acta. Arith. **88** (1999), no. 1, 67–98.

[34] P. Tomas, *A restriction theorem for the Fourier transform*, Bull. Amer. Math. Soc. **81** (1975), 477–478.

[35] P. Varnavides, *On certain sets of positive density,* J. London Math. Soc. **34** (1959), 358–360.

[36] R. Vaughan, *The Hardy-Littlewood method, 2nd ed,* Cambridge Tracts in Math. **125**, CUP 1997.

[37] D.R. Ward *Some series involving Euler's function,* J. London Math. Soc. **2** (1927), 210–214.

[38] A. Zygmund, *Trigonometric series,* 2nd ed. Vols I, II. CUP 1959.

Ben GREEN
School of Mathematics
University of Bristol
Bristol BS8 1TW
England
E-mail : `b.j.green@bristol.ac.uk`
URL: `http://www.maths.bris.ac.uk/~mabjg`

Terence TAO
Department of Mathematics
University of California at Los Angeles
Los Angeles CA 90095
E-mail : `tao@math.ucla.edu`
URL: `http://www.math.ucla.edu/~tao`